Data Analysis and Manifold Learning
Lecture 2: Properties of Symmetric Matrices and
Examples

Radu Horaud
INRIA Grenoble Rhone-Alpes, France
Radu.Horaud@inrialpes.fr
http://perception.inrialpes.fr/
Outline of Lecture 2

- Basic definitions, eigen decomposition, LU and Cholesky matrix factorizations;
- Spectral decomposition, powers, inverse, exponential;
- Geometric interpretation;
- The Raleigh-Ritz theorem and extensions;
- Computing eigenvalues and eigenvectors in practice: power method, inverse power method, and shifted inverse power method;
Material for This Lecture

- Software: http://www.caam.rice.edu/software/ARPACK/ written in Fortran77!
Some Basic Definitions

- Symmetry of a $D \times D$ matrix: $A = A^\top$
- Eigen decomposition: $A = U\Lambda U^\top$ with the properties:
  - $UU^\top = U^\top U = I_D$
  - $\det(U) = \pm 1$
  - All the eigenvalues are real numbers:
    
    $$\lambda_{\min} = \lambda_1 \leq \ldots \leq \lambda_i \leq \ldots \leq \lambda_D = \lambda_{\max}$$

- $A$ is referred to as a *real symmetric matrix*;
- If $\lambda_1 \geq 0$ then it is a *positive semi-definite symmetric matrix*
- If $\lambda_1 > 0$ then it is a *positive definite symmetric matrix*
- Symmetric matrices are *nondefective*: the algebraic multiplicity of each eigenvalue is equal to its geometric multiplicity.
Spectral Decomposition, Deflation, Powers, Exponential

- A symmetric matrix can be written as $A = \sum_{i=1}^{D} \lambda_i u_i u_i^\top$ where $\{u_i\}_{i=1}^{D}$ are the eigenvectors or, equivalently, the column vectors of $U$.
- The transformation $\tilde{A} = A - \lambda_k u_k u_k^\top$ is known as a deflation.
- Note that $\tilde{A} u_k = 0$.
- $A^2 = U \Lambda U^\top U \Lambda U^\top = U \Lambda^2 U^\top$
- More generally: $A^k = U \Lambda^k U^\top$
- The matrices $A, A^2, \ldots, A^k$ have the same eigenvectors $\{u_i\}$ and eigenvalues $\lambda, \lambda^2, \ldots, \lambda^k$.
- Matrix exponential: $e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$
- We have: $e^A = U \text{Diag}[e^{\lambda_1} \ldots e^{\lambda_i} \ldots e^{\lambda_D}] U^\top$
- Hence, matrix $e^A$ has the same eigenvectors as $A$ and eigenvalues $e^\lambda$. 

Radu Horaud  Data Analysis and Manifold Learning; Lecture 2
The Inverse of a Symmetric Matrix

- The case of a non-singular symmetric matrix:
  - $A^{-1} = U\Lambda^{-1}U^\top$.
  - Spectral decomposition: $A^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} u_i u_i^\top$.
  - The matrices $A^{-1}, A^{-2}, \ldots, A^{-k}$ have eigenvectors $u_i$ and eigenvalues $\lambda_i^{-1}, \lambda_i^{-2}, \ldots, \lambda_i^{-k}$. 

Radu Horaud
Data Analysis and Manifold Learning; Lecture 2
The Pseudoinverse of a Singular Symmetric Matrix

- If a symmetric matrix is singular, it has as eigenvalue $\lambda = 0$ with multiplicity $m > 0$.
- We rearrange the eigenvalue-eigenvector pairs and we retain the non-zero pairs, such that:

$$A = \tilde{U}\tilde{\Lambda}\tilde{U}^\top$$

with: $\tilde{\Lambda} = \text{Diag}[\lambda_1 \ldots \lambda_{D-m}]$ and $\tilde{U} = [u_1 \ldots u_{D-m}]$.
- $\tilde{U}$ is a $D \times (D - m)$ matrix whose columns are orthogonal. Notice that $\tilde{U}^\top\tilde{U} = I_{D-m}$ but $\tilde{U}\tilde{U}^\top \neq I_D$!
- The Moore-Penrose pseudoinverse:

$$A^\dagger = \tilde{U}\text{Diag}[\lambda_1^{-1} \ldots \lambda_{D-m}^{-1}]\tilde{U}^\top$$
The Choleski Factorization

- We consider the case of positive definite symmetric matrices. They can be written as $A = BB^\top$ but the choice of $B$ is not unique.
- Any such matrix can be decomposed as: $A = LL^\top$ with $L$ being a low-triangular matrix with nonnegative diagonal entries. This decomposition is unique.
- Complexity of Choleski decomposition algorithms for a $D \times D$ non singular matrix: $D^3$ FLOPS. This is twice more efficient than the LU decomposition.
- Let $Ax = b$. No matrix inversion needed to solve it! This can be rewritten as:

$$\begin{align*}
Ly &= b \\
L^\top x &= y
\end{align*}$$
Matrix Norms

- The Frobenius norm:

\[ \| A \|_F^2 = \text{tr}(A^\top A) = \text{tr}(U\Lambda^2U^\top) = \text{tr}(\Lambda^2) = \sum_{i=1}^{D} \lambda_i^2 \]

- The spectral norm:

\[ \max_{\| v \|} \frac{\| Av \|}{\| v \|} = \left( \max_{\| v \|} \frac{v^\top A^\top A v}{v^\top v} \right)^{1/2} = \lambda_{\text{max}} \]

(see the Rayleigh-Ritz theorem below)
Geometric Interpretation

- Consider a positive definite symmetric matrix; In this case all the eigenvalues are strictly positive.
- Quadratic form for any vector $x \neq 0$:

$$x^\top Ax = (U^\top x)^\top \Lambda (U^\top x) = \sum_{i=1}^{D} \lambda_i (u_i^\top x)^2$$

- Let’s transform the data into another coordinate frame: $z = U^\top x$; we obtain: $x^\top Ax = z^\top \Lambda z$.

$$z^\top \Lambda z = (z_1/\lambda_1^{-1/2})^2 + \ldots (z_D/\lambda_D^{-1/2})^2 = C$$

- This is an ellipsoid with axes $u_1 \ldots u_D$ and with half eccentricities $\lambda_1^{-1/2} \ldots \lambda_D^{-1/2}$ (Remember PCA...).
The Rayleigh-Ritz Theorem

Theorem

(Rayleigh-Ritz). Let $A$ be a symmetric matrix with ordered eigenvalues, then:

$\lambda_1 x^\top x \leq x^\top A x \leq \lambda_D x^\top x \ \forall x$

$\lambda_{\text{max}} = \lambda_D = \max_{x \neq 0} \frac{x^\top A x}{x^\top x} = \max_{x^\top x = 1} x^\top A x$

$\lambda_{\text{min}} = \lambda_1 = \min_{x \neq 0} \frac{x^\top A x}{x^\top x} = \min_{x^\top x = 1} x^\top A x$
Proof of the Rayleigh-Ritz Theorem

- From the eigendecomposition: \( \mathbf{x}^\top \mathbf{A} \mathbf{x} = \sum_{i=1}^{D} \lambda_i \left( (\mathbf{U}^\top \mathbf{x})_i \right)^2 \)
- Notice that: \( \sum_{i=1}^{D} \left( (\mathbf{U}^\top \mathbf{x})_i \right)^2 = \| \mathbf{U}^\top \mathbf{x} \|^2 = \| \mathbf{x} \|^2 = \mathbf{x}^\top \mathbf{x} \)
- Using the fact that the eigenvalues can be ordered, we get the first part of the theorem.
- By dividing we obtain: \( \lambda_{\text{min}} \leq \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \leq \lambda_{\text{max}}, \ (\mathbf{x} \neq 0) \)
- with equalities when \( \mathbf{x} \) is a \( \lambda_1 \) or \( \lambda_D \) eigenvector.
- We have: \( \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = (\mathbf{x}^\top / \sqrt{\mathbf{x}^\top \mathbf{x}}) \mathbf{A} (\mathbf{x} / \sqrt{\mathbf{x}^\top \mathbf{x}}) \) and hence the minimization/maximization of the Raleigh quotient is equivalent to:

\[
\begin{cases}
\max \mathbf{x} \ \mathbf{x}^\top \mathbf{A} \mathbf{x} \\
\mathbf{x}^\top \mathbf{x} = 1 
\end{cases}
\]
What About the Remaining Eigenvalues and Eigenvectors?

- Let’s restrict $x$ to be orthogonal to the smallest eigenvector $u_1$, i.e, $u_1^\top x = 0$:

- $x^\top Ax = \sum_{i=2}^{D} \lambda_i ((U^\top x)_i)^2 \geq \lambda_2 x^\top x$

- with equality when $x = u_2$

- Therefore we obtain:

$$
\lambda_2 = \min_{x^\top x = 1, x^\top u_1 = 0} x^\top Ax
$$

$$
\lambda_{D-1} = \max_{x^\top x = 1, x^\top u_D = 0} x^\top Ax
$$
Computing Eigenvalues and Eigenvectors in Practice

- The *power method* estimates the largest eigenvalue/eigenvector pair or an *eigenpair*.
- The *power method + deflation* estimates the second largest eigenpair, etc.
- The *inverse power method* estimates the smallest eigenpair.
- The *shifted inverse power method* allows to obtain intermediate eigenpairs.
- The Lanczos method is an adaptation of the power method. It is very useful for large and sparse matrices. It is used by the ARPACK package.
The Power Method

- Input: A symmetric matrix $A$ and a random vector $x_0$.
- At each iteration $k$:
  1. Normalize $y_k = \frac{x_k}{\|x_k\|}$ and
  2. $x_{k+1} = Ay_k$.
- Check for convergence: $\|y_{k+1} - y_k\| < \varepsilon$
- Output: $u_D = y_{k+1}$ and $\lambda_D = y_{k+1}^\top Ay_{k+1}$
Justification of the Power Method

Let \( x_0 = \sum_{i=1}^{D} \alpha_i u_i \) hence we obtain after the first iteration:
\[
x_1 = A x_0 = \sum_{i=1}^{D} \alpha_i \lambda_i u_i
\]

Normalize this vector: \( y_1 = \frac{1}{\beta_1} \sum_{i=1}^{D} \alpha_i \lambda_i u_i \)

More generally: \( y_{k+1} = \frac{1}{\beta_1 \ldots \beta_{k+1}} \sum_{i=1}^{D} \alpha_i \lambda_i^{k+1} u_i \)

At the limit this vector becomes the “largest” eigenvector:
\[
y_\infty = \lim_{k \to \infty} \frac{\alpha_D \lambda_D^{k+1}}{\beta_1 \ldots \beta_{k+1}} \left( \sum_{i=1}^{D-1} \frac{\alpha_i}{\alpha_D} \frac{\lambda_i^{k+1}}{\lambda_D^{k+1}} u_i + u_D \right) = u_D
\]
\[
\lambda_D = y_\infty^\top A y_\infty
\]
The Power Method with Deflation

- Consider the matrix \( \tilde{A} = A - \lambda_D u_D u_D^T \)
- Notice that \((0, u_D)\) is an eigenpair of \( \tilde{A} \) and that the remaining eigenpairs remain unchanged (refer to the spectral decomposition of \( A \) and to the fact that eigenvectors corresponding to distinct eigenvalues are orthogonal).
- It follows that the second largest eigenpair \((\lambda_{D-1}, u_{D-1})\) of \( A \) becomes the largest eigenpair of \( \tilde{A} \).
- The power method can now be applied to \( \tilde{A} \), etc.
The Inverse Power Method

- The smallest eigenvector-eigenvalue pair \((u_1, \lambda_1)\) of \(A\) corresponds to the largest eigenvector-eigenvalue pair \((u_1, \lambda_1^{-1})\) of \(A^{-1}\).
- The \(k\)-th iteration of the power method becomes:
  \[ x_{k+1} = A^{-1} y_k \]
- which can be written as:
  \[ A \cdot x_{k+1} = y_k \]
- This can be solved using the Choleski factorization \(A = LL^\top\):
  \[ \begin{cases} Lz = y_k \\ L^\top x_{k+1} = z \end{cases} \]
The Shifted Inverse Power Method

- Let’s consider the matrix $B = A - \alpha I$ as well as an eigenpair $Au = \lambda u$.

- $(\lambda - \alpha, u)$ becomes an eigenpair of $B$, indeed:

$$Bu = (A - \alpha I)u = (\lambda - \alpha)u$$

and hence $B$ is a real symmetric matrix with eigenpairs $(\lambda_1 - \alpha, u_1), \ldots (\lambda_i - \alpha, u_i), \ldots (\lambda_D - \alpha, u_D)$.

- If $\alpha > 0$ is chosen such that $|\lambda_j - \alpha| \ll |\lambda_i - \alpha| \ \forall i \neq j$ then $\lambda_j - \alpha$ becomes the smallest (in magnitude) eigenvalue.

- The inverse power method (in conjunction with the $LU$ decomposition of $B$) can be used to estimate the eigenpair $(\lambda_j - \alpha, u_j)$. 

Radu Horaud

Data Analysis and Manifold Learning; Lecture 2